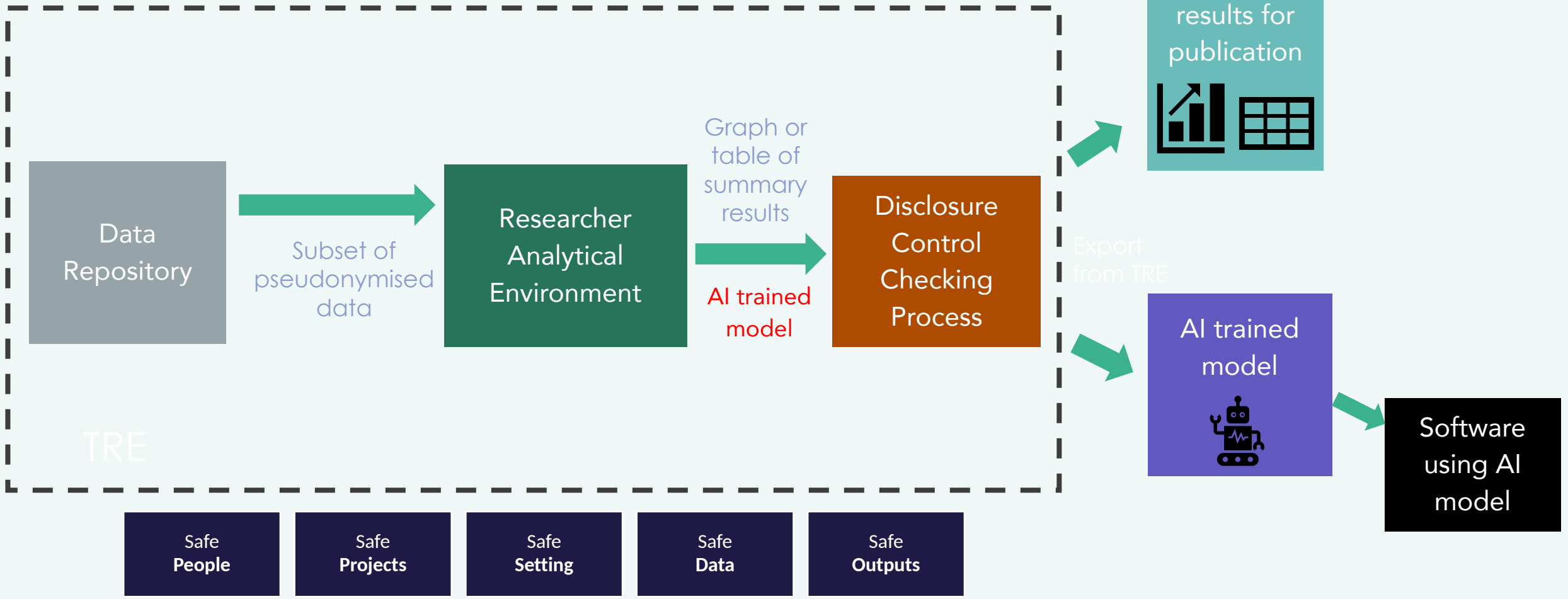
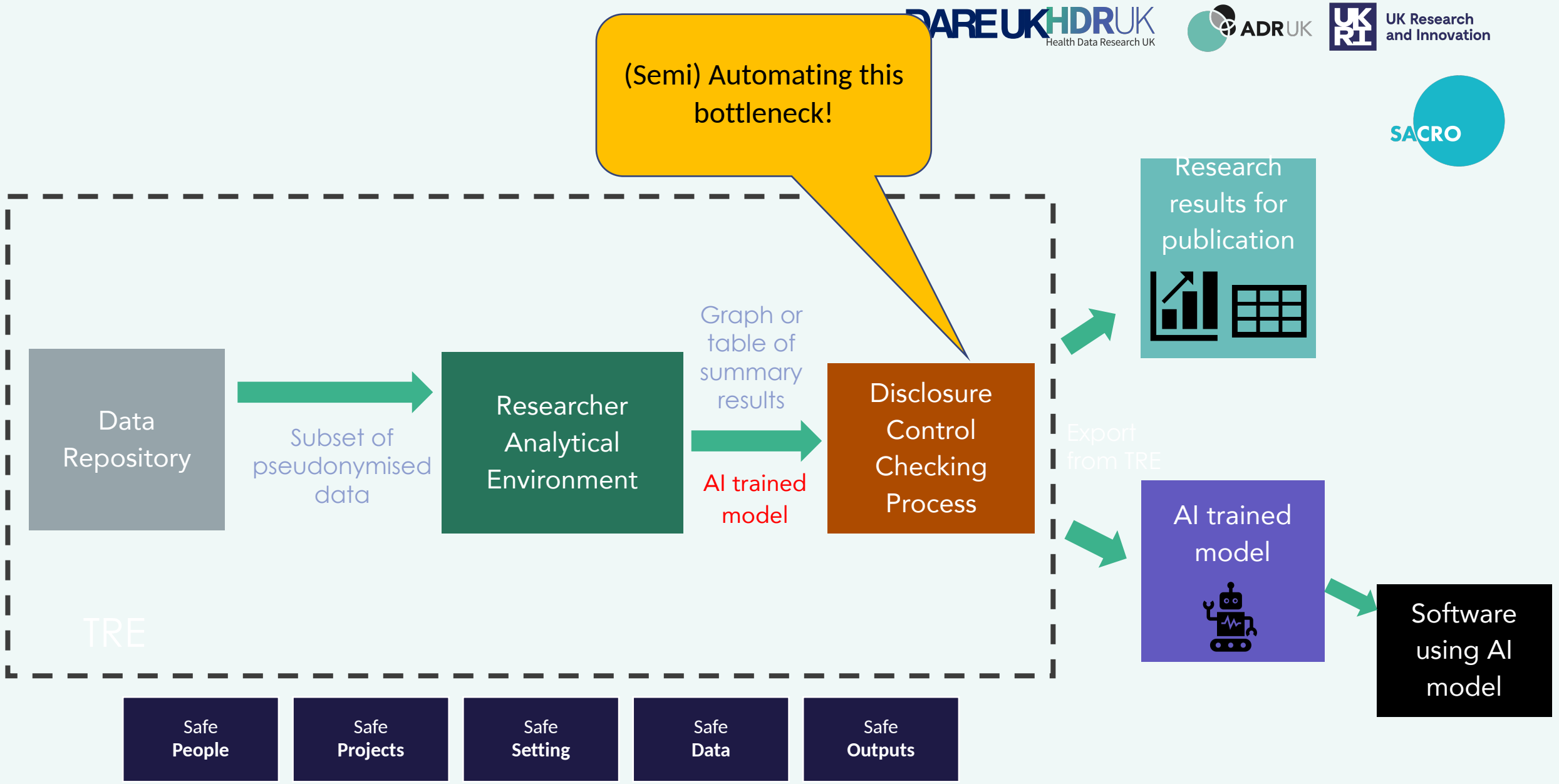
SACRO: Semi-Automated Checking of Research Outputs

Professor Jim Smith,

University of the West of England



SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023



SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Who are we? (alphabetically)

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Who are we? (alphabetically)

Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Who are we? (alphabetically)

Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

Public Data Bodies

- Health Data Research UK
- NHS Scotland
- Public Health Scotland
- Research Data Scotland

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Who are we? (alphabetically)

Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

Public Data Bodies

- Health Data Research UK
- NHS Scotland
- Public Health Scotland
- Research Data Scotland

TREs

- DASH (Aberdeen/Grampian)
- DataLoch (Edinburgh)
- HIC (Dundee)
- eDRIS (Public Health Scot)
- OpenSafely (Oxford)

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Who are we? (alphabetically)

Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

Public Data Bodies

- Health Data Research UK
- NHS Scotland
- Public Health Scotland
- Research Data Scotland

TREs

- DASH (Aberdeen/Grampian)
- DataLoch (Edinburgh)
- HIC (Dundee)
- eDRIS (Public Health Scot)
- OpenSafely (Oxford)

External steering group:

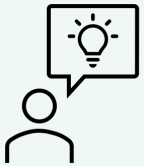
UK: Office for National Statistics, ESRC, DARE
Global: Eurostat, SDC-GESIS, ICPSR (US)

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation

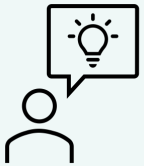
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation



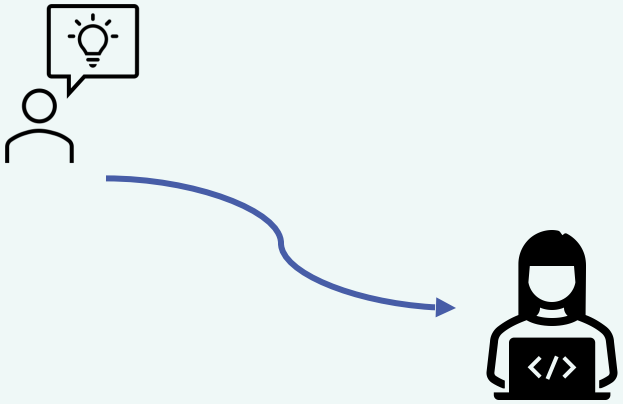
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation



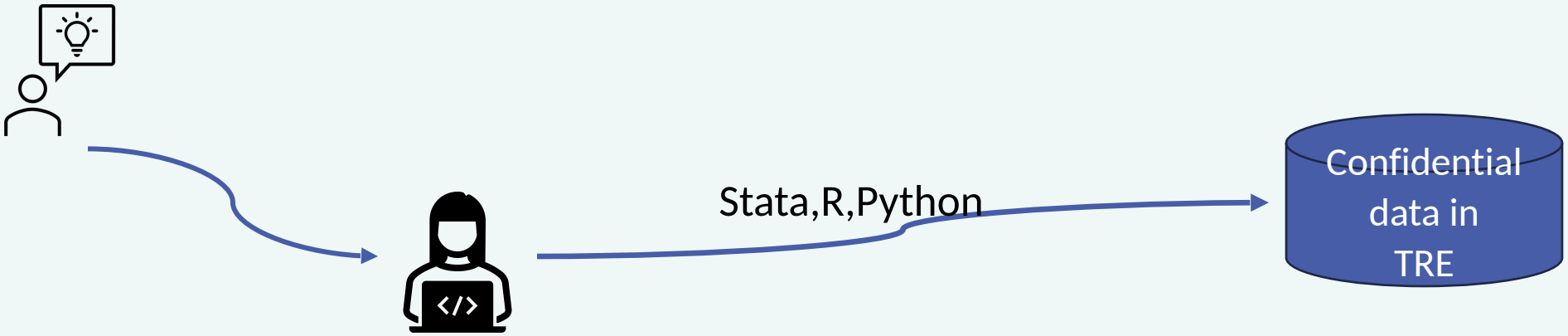
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation



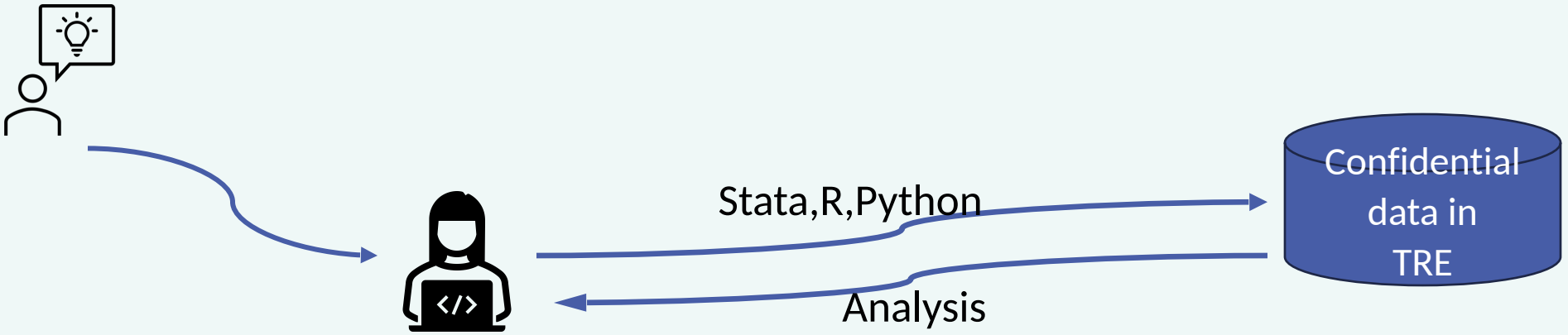
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation



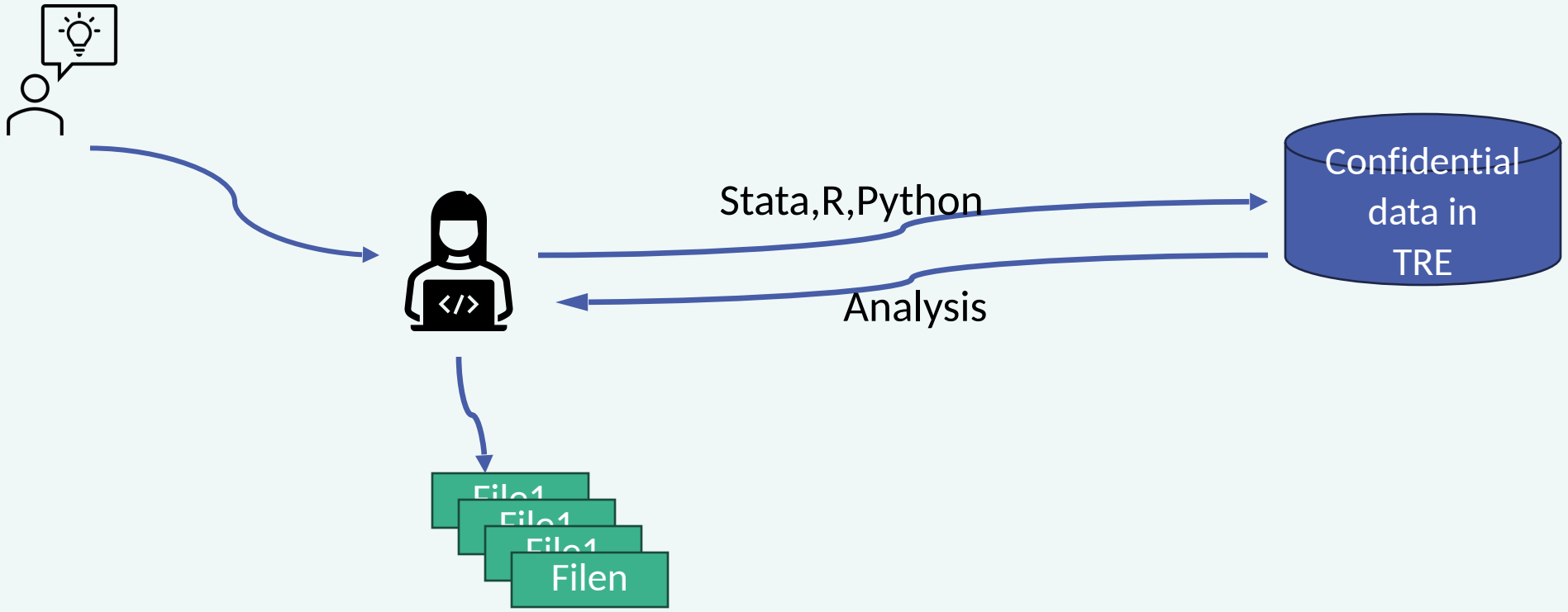
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation



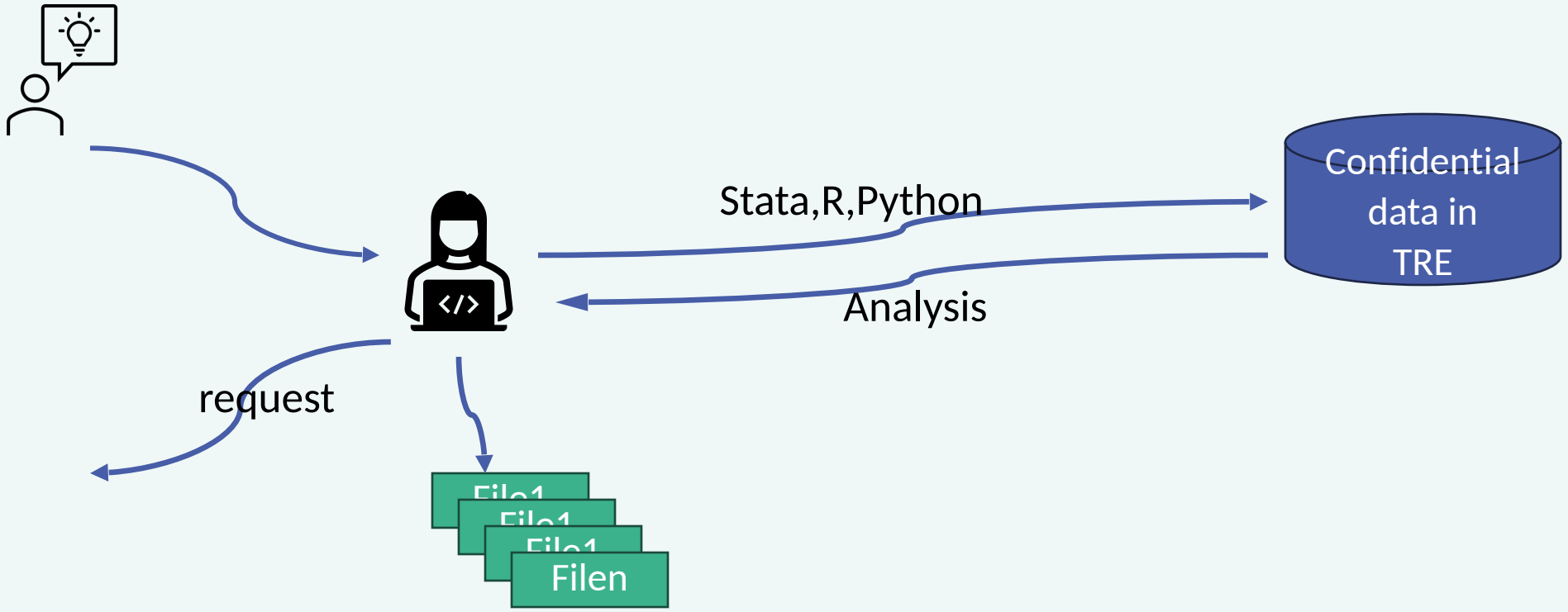
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation



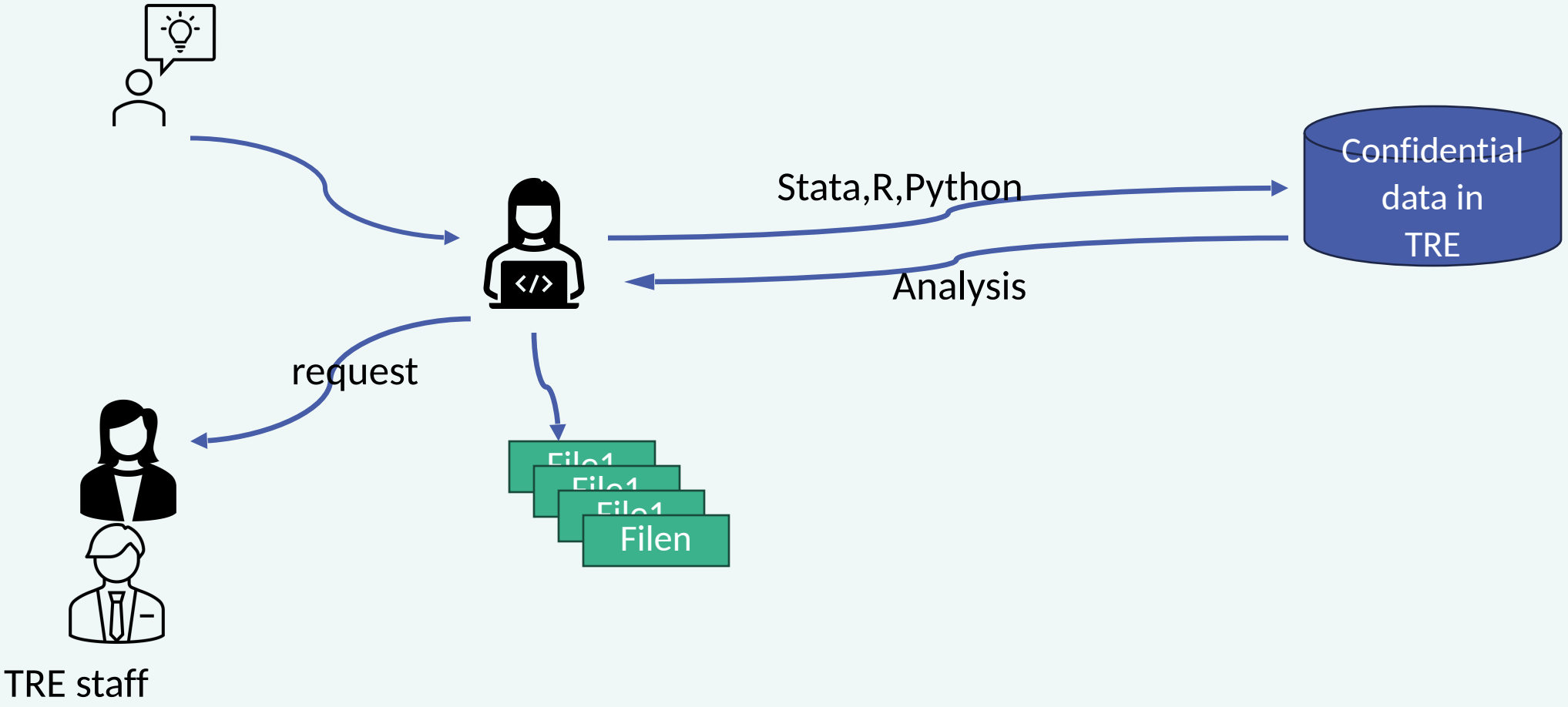
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation



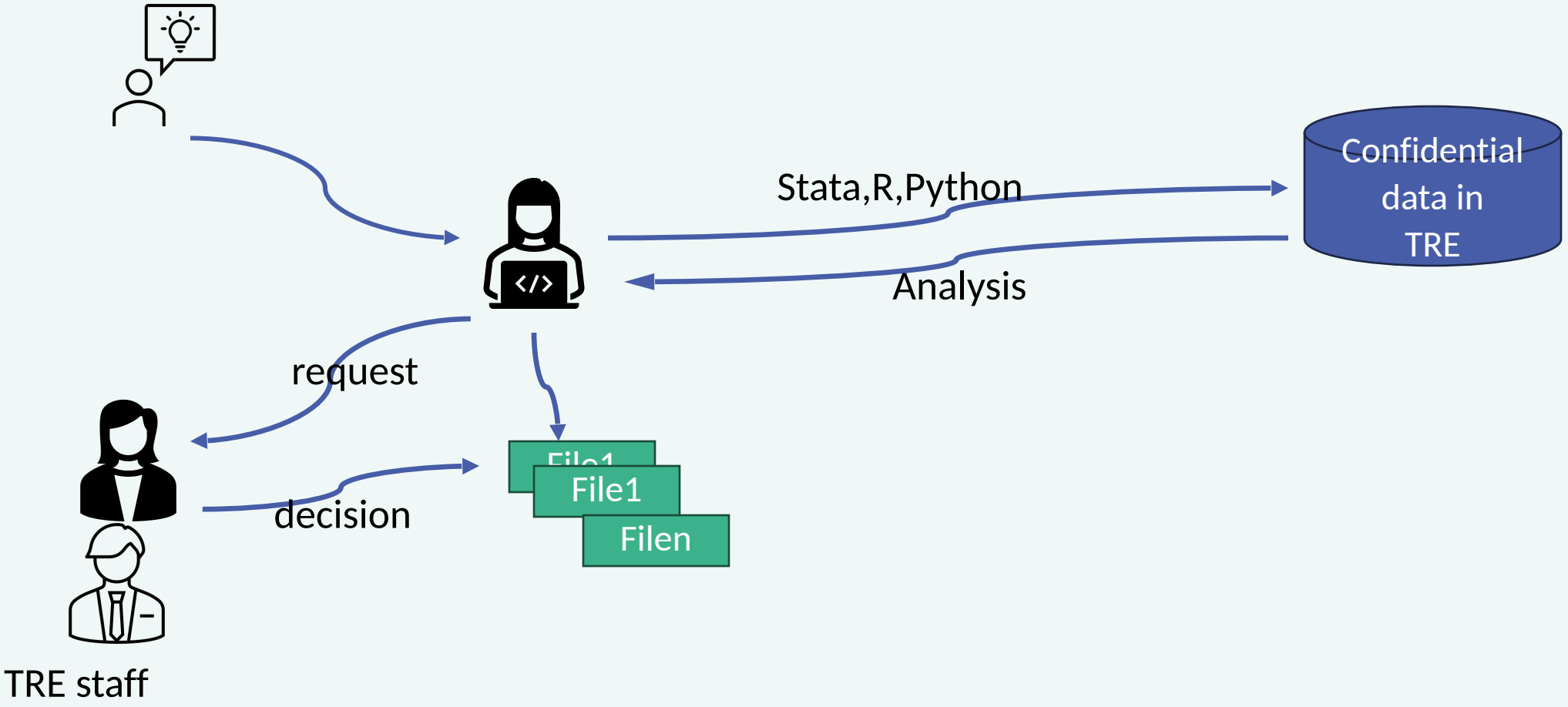
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation



SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

The current situation

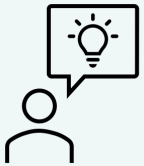


SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell

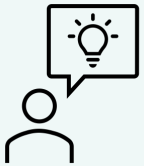
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

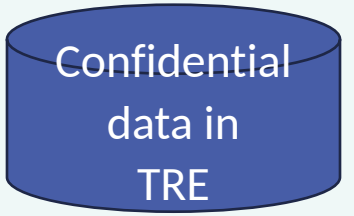
SACRO in a nutshell



SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

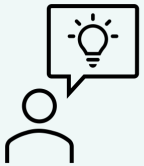
SACRO in a nutshell





SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell

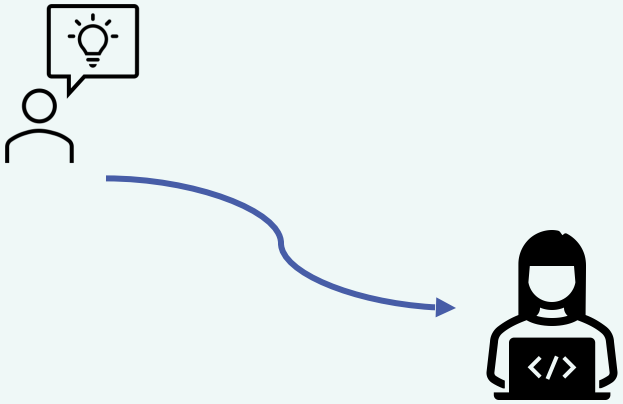


Confidential data in TRE

TRE risk appetite

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell

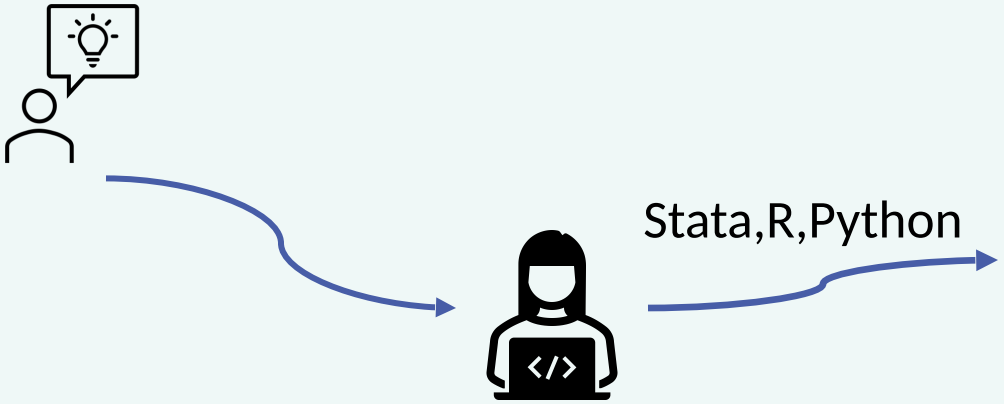


Confidential data in TRE

TRE risk appetite

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell

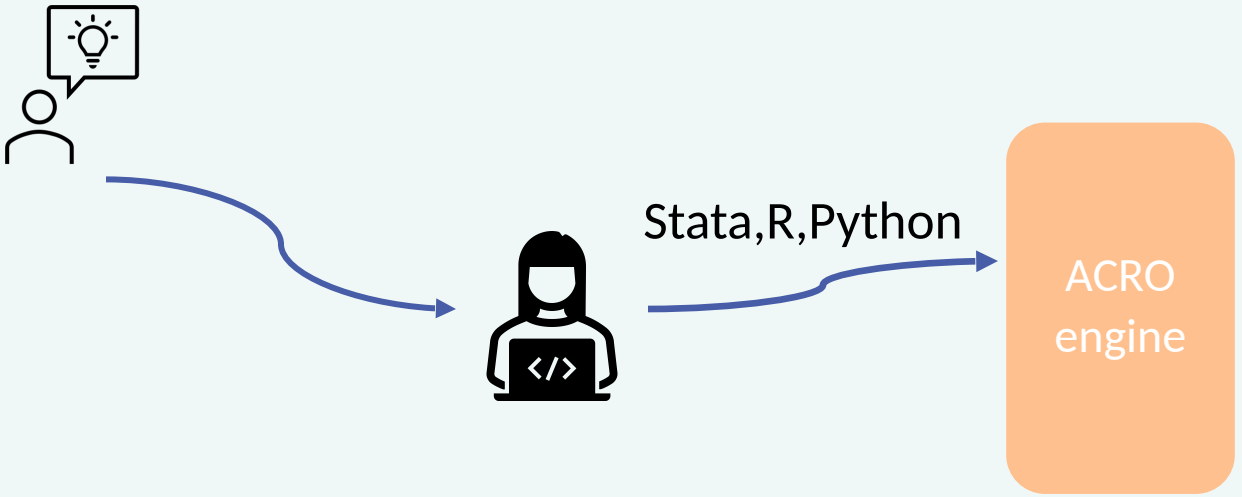


Confidential data in TRE

TRE risk appetite

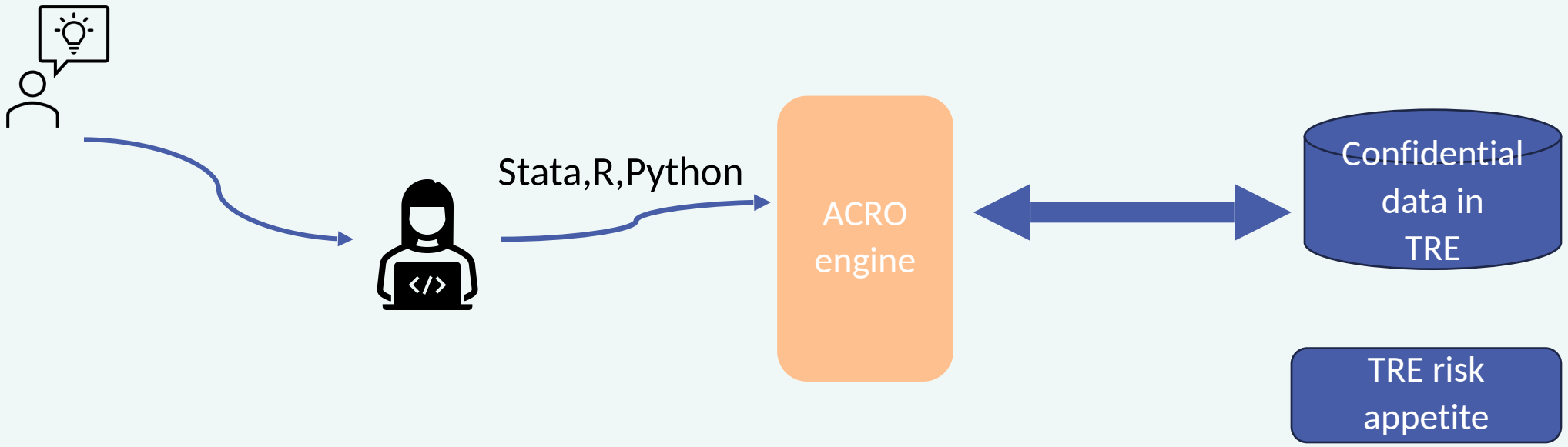
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell



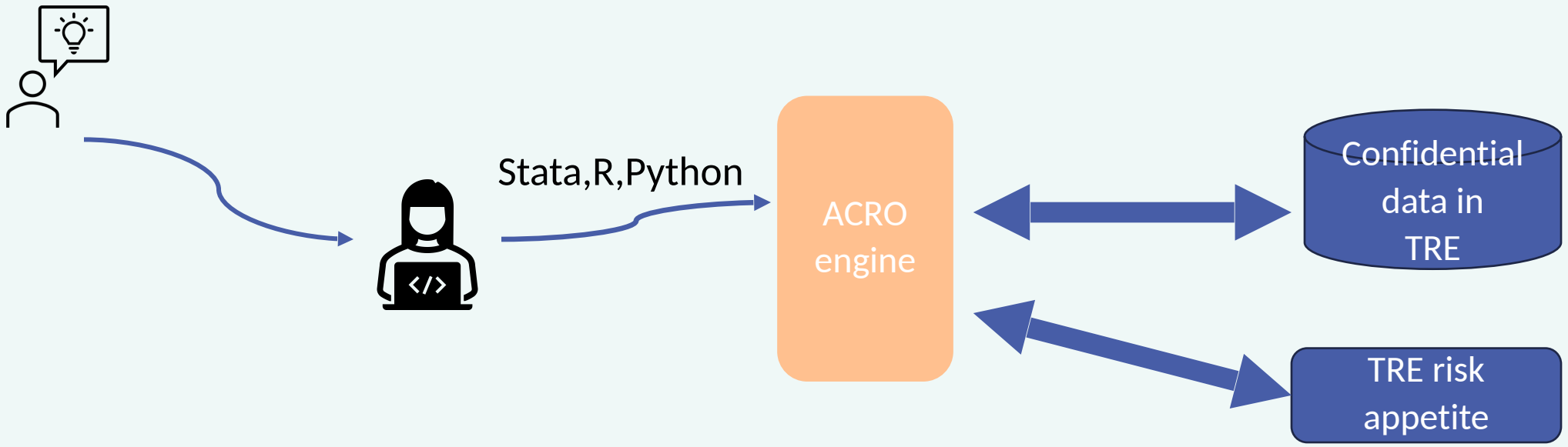
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell



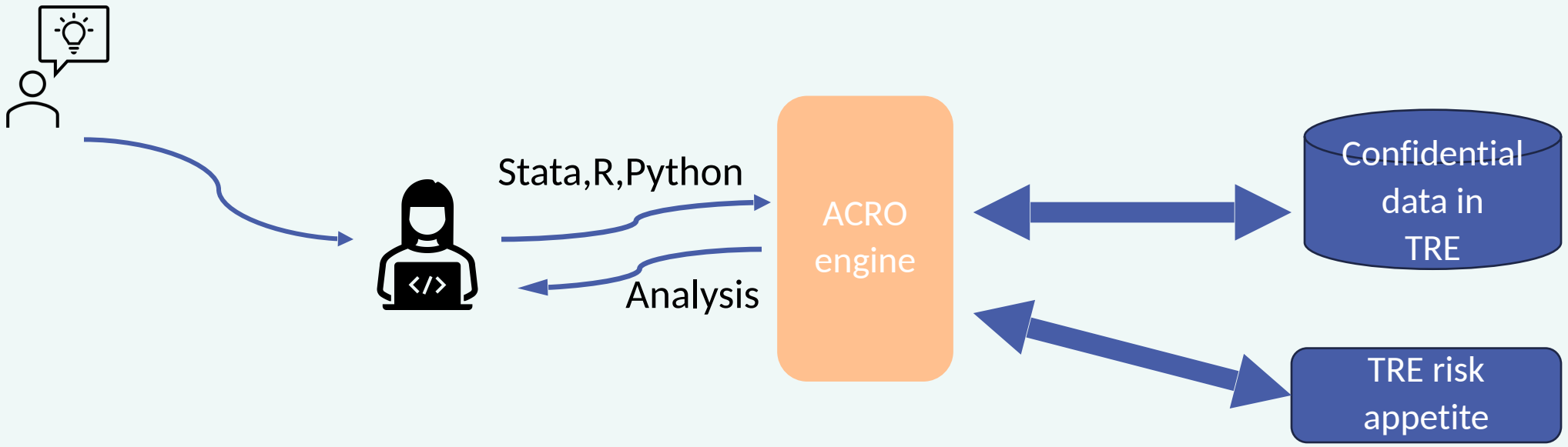
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell



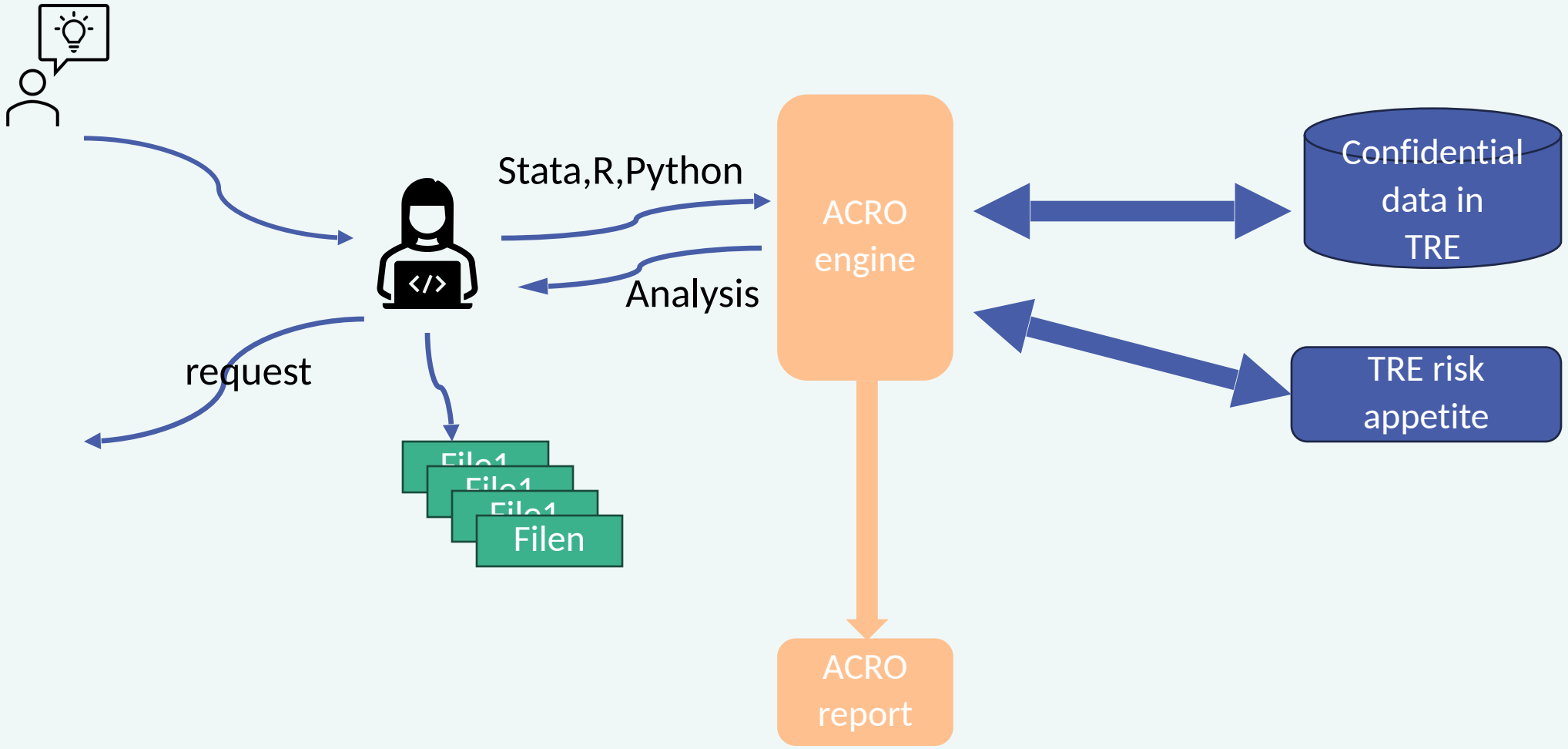
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell



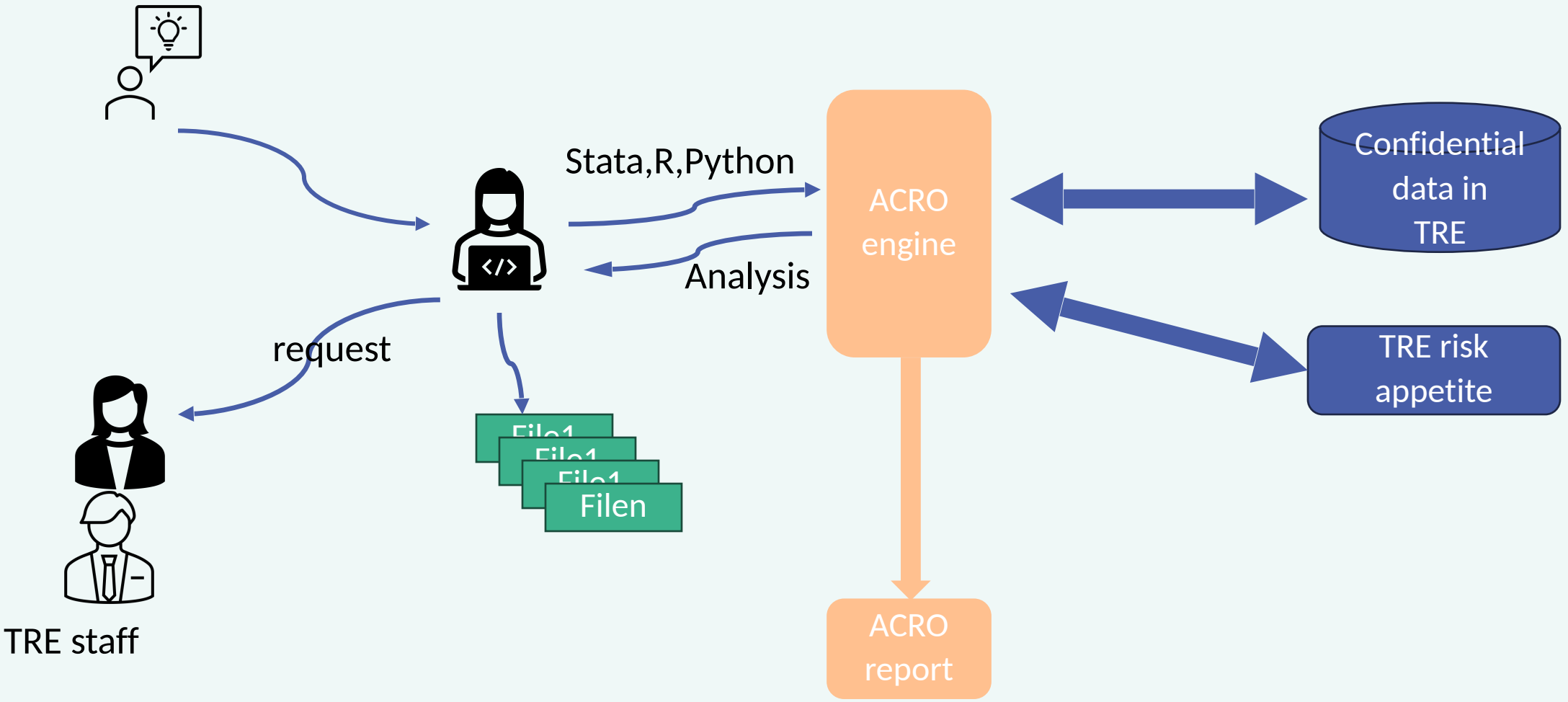
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell



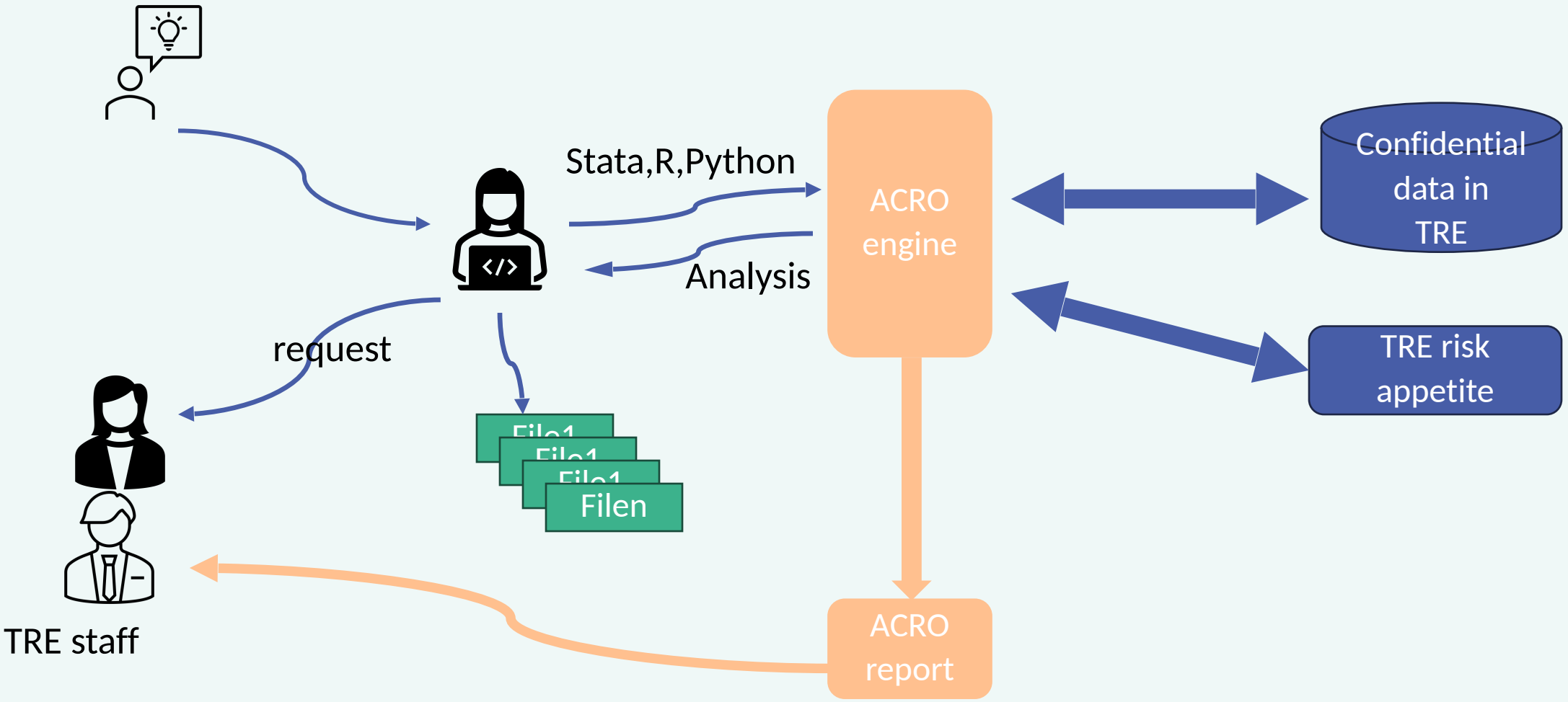
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell



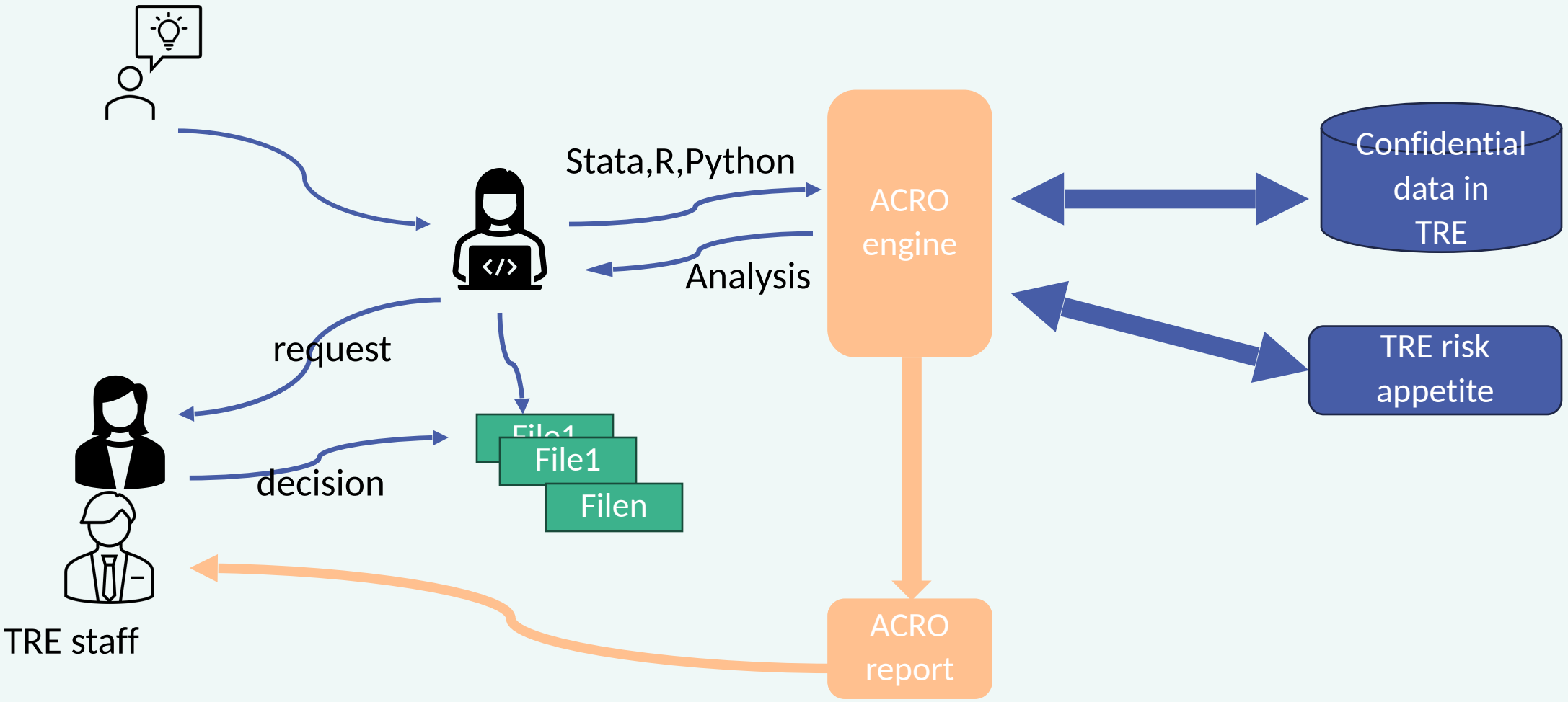
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell



SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO in a nutshell



SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Similar for Machine Learning Models

Except that we :

Similar for Machine Learning Models

Except that we :

- Run a range of “inference” attacks”

Similar for Machine Learning Models

Except that we :

- Run a range of “inference” attacks”
- Aim to support more ‘user journeys’

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Similar for Machine Learning Models

Except that we :

- Run a range of “inference” attacks”
- Aim to support more ‘user journeys’
- Don’t have a set of ‘tried and trusted’ guidelines to work with

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

ACRO: analytic commands

ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- **acro.crosstab()**
- **acro.pivot_table**()

Using the same api as statsmodels, adding checks for DoF:

ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- **acro.crosstab()**
- **acro.pivot_table**()

Using the same api as statsmodels, adding checks for DoF:

- **acro.logit() , acro.logitr()**

ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- **acro.crosstab()**
- **acro.pivot_table**()

Using the same api as statsmodels, adding checks for DoF:

- **acro.logit() , acro.logitr()**
- **acro.ols() , acro.olsr()**

ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- **acro.crosstab()**
- **acro.pivot_table**()

Using the same api as statsmodels, adding checks for DoF:

- **acro.logit() , acro.logitr()**
- **acro.ols() , acro.olsr()**
- **acro.probit() , acro.probitr()**

ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- **acro.crosstab()**
- **acro.pivot_table**()

Using the same api as statsmodels, adding checks for DoF:

- **acro.logit() , acro.logitr()**
- **acro.ols() , acro.olsr()**
- **acro.probit() , acro.probitr()**

R versions are prefixed by “acro_”

ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- **acro.crosstab()**
- **acro.pivot_table**()

Using the same api as statsmodels, adding checks for DoF:

- **acro.logit() , acro.logitr()**
- **acro.ols() , acro.olsr()**
- **acro.probit() , acro.probitr()**

R versions are prefixed by “acro_”

- Also support R’s built in ‘**table**’ command

ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- **acro.crosstab()**
- **acro.pivot_table**()

Using the same api as statsmodels, adding checks for DoF:

- **acro.logit() , acro.logitr()**
- **acro.ols() , acro.olsr()**
- **acro.probit() , acro.probitr()**

R versions are prefixed by “acro_”

- Also support R’s built in ‘**table**’ command
- Stata versions prefixed by “acro ”
  - Code currently captures and translates commands: **table, regress, probit, logit**

ACRO: user commands for session management

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

ACRO: user commands for session management

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

ACRO: user commands for session management

- acro = ACRO()

ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]

ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()

ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)

ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)

ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)

ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)
- acro.add_comment(key, text)

ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)
- acro.add_comment(key, text)
- acro.custom_output(filename, description)

ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)
- acro.add_comment(key, text)
- acro.custom_output(filename, description)
- acro.finalise(output_directory,format)

Example: what happens if researcher requests a disclosive table

```
acro.suppress = False

safe_table = acro.crosstab(df.recommend, df.parents)
print(safe_table)
```

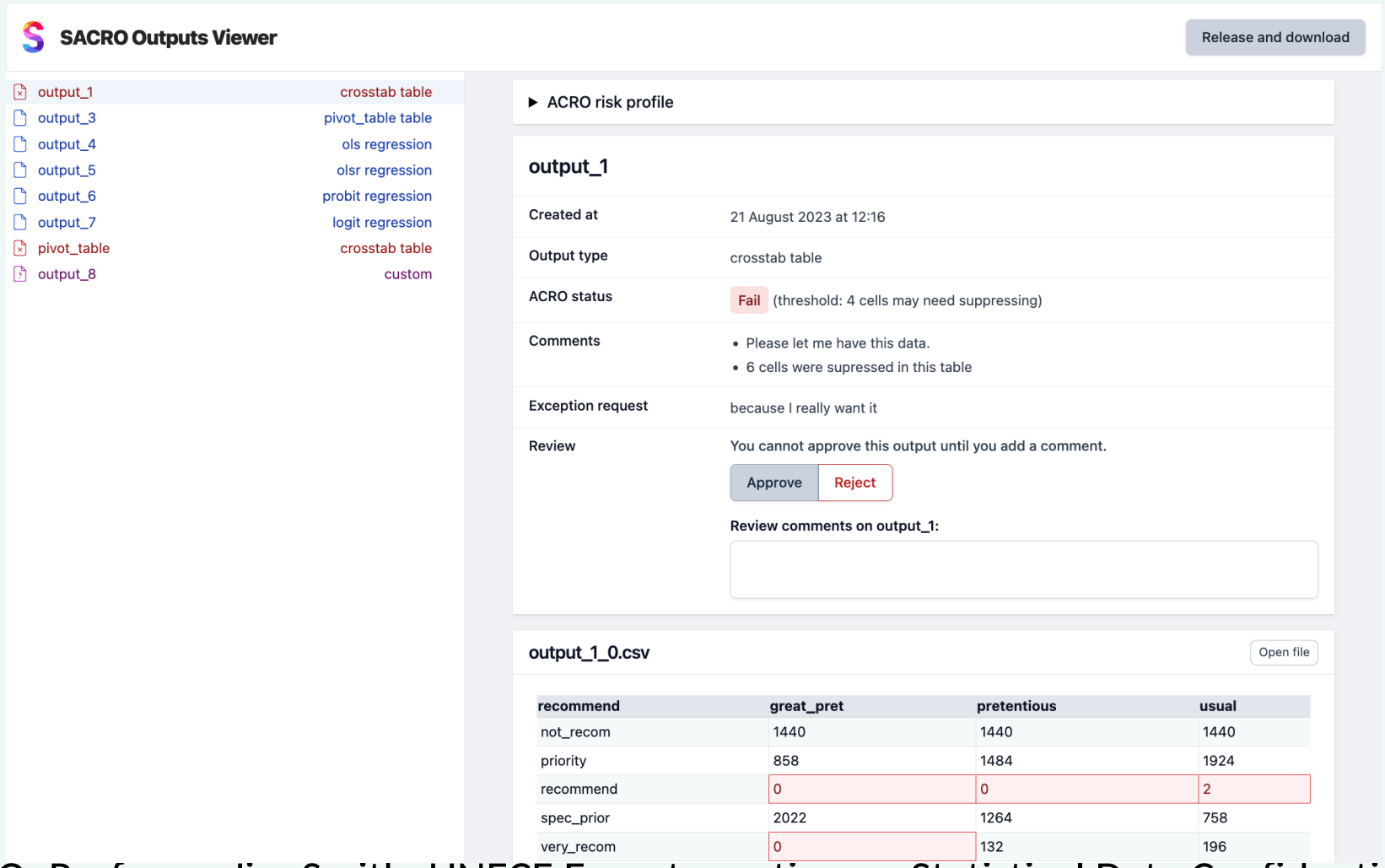
```
INFO:acro:get_summary(): fail; threshold: 4 cells may need suppressing;
INFO:acro:outcome_df:
-------------------------------------------------|
parents     |great_pret    |pretentious   |usual       |
recommend   |              |              |            |
-------------------------------------------------|
not_recom   |           ok |           ok |          ok|
priority    |           ok |           ok |          ok|
recommend   | threshold;   | threshold;   | threshold; |
spec_prior  |           ok |           ok |          ok|
very_recom  | threshold;   |           ok |          ok|
-------------------------------------------------|

INFO:acro:records:add(): output_1
```

parents	great_pret	pretentious	usual
recommend			
not_recom	1440	1440	1440
priority	858	1484	1924
recommend	0	0	2
spec_prior	2022	1264	758
very_recom	0	132	196

23

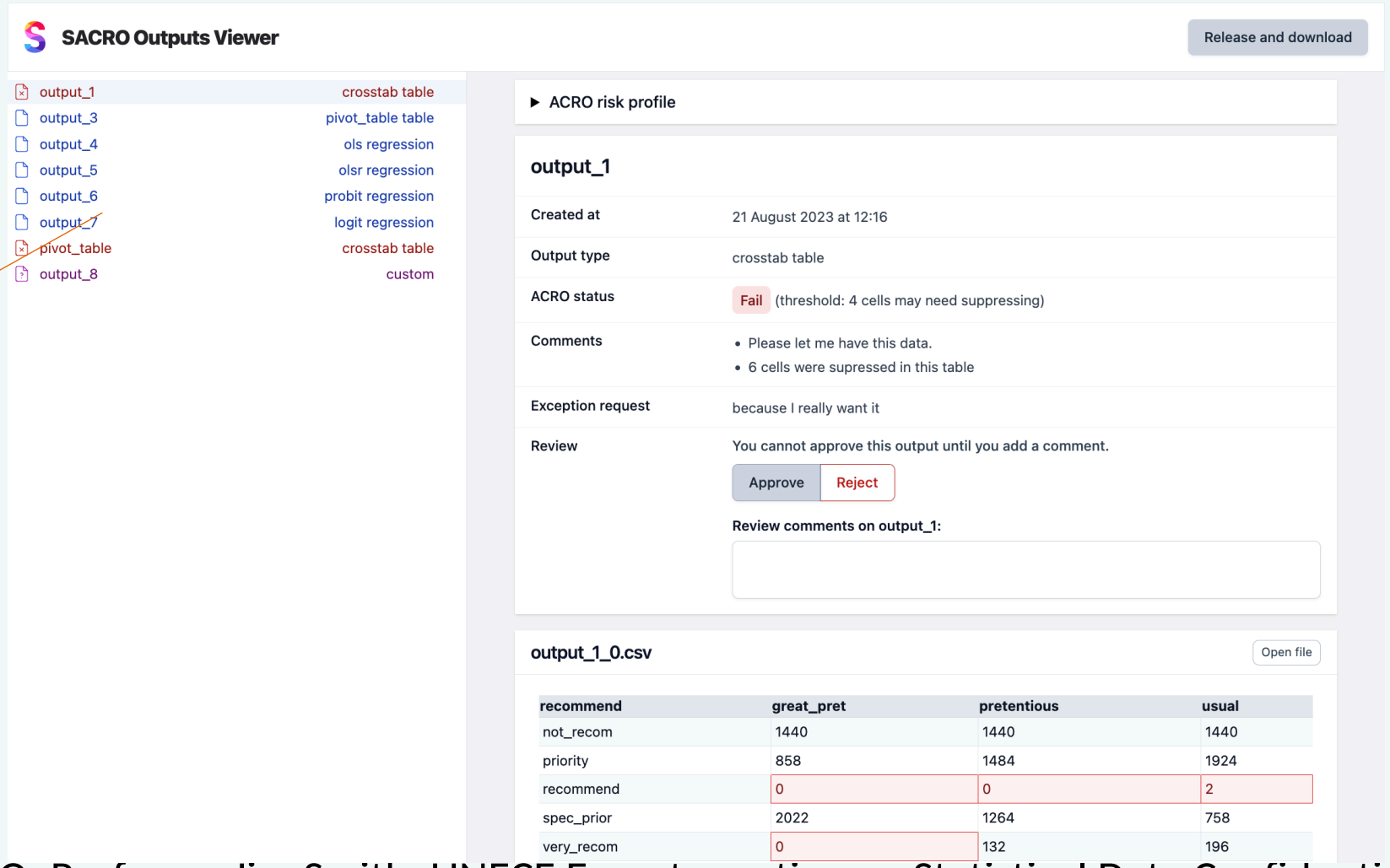
SACRO viewer for TRE output checkers



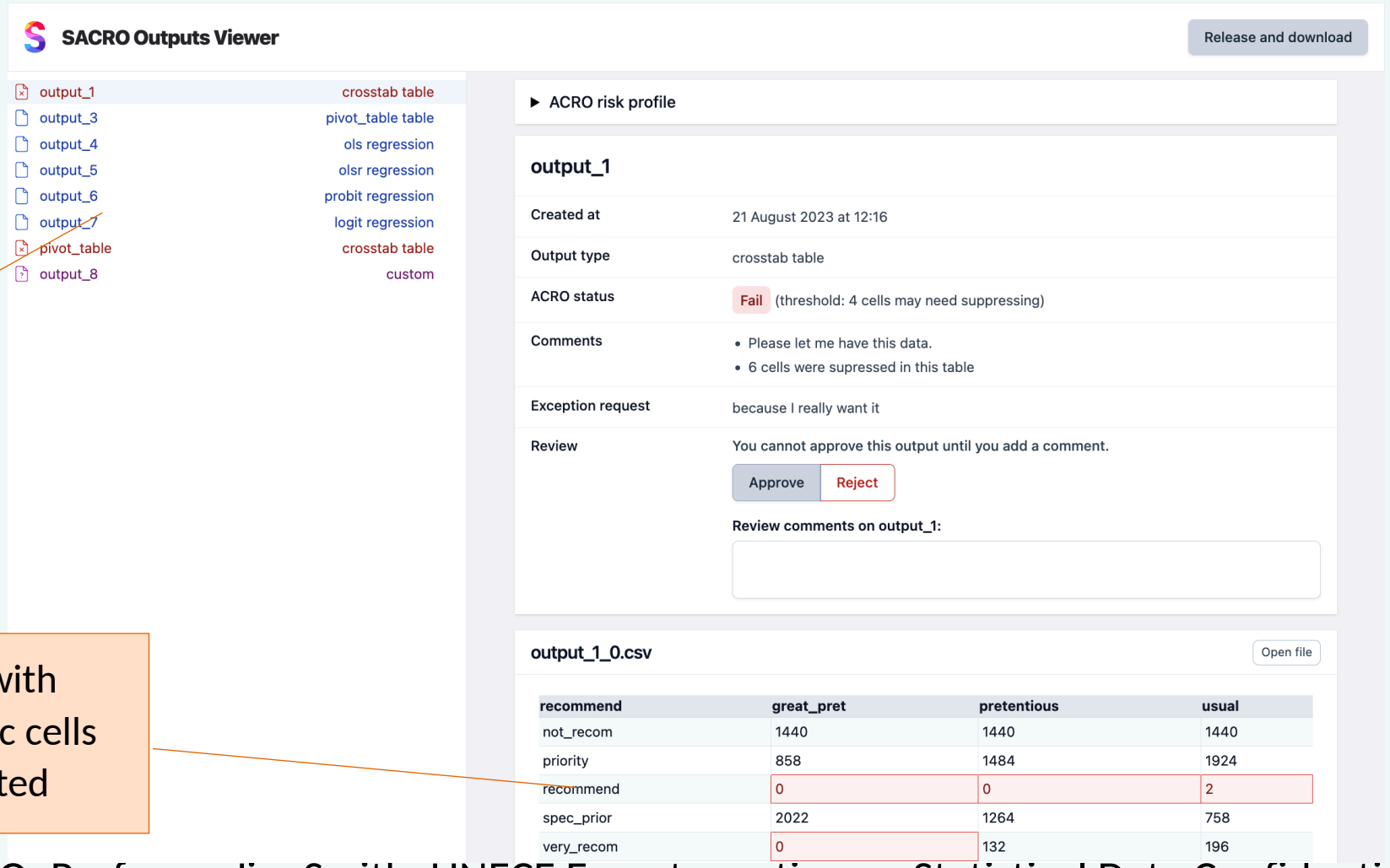
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO viewer for TRE output checkers

Outputs to review



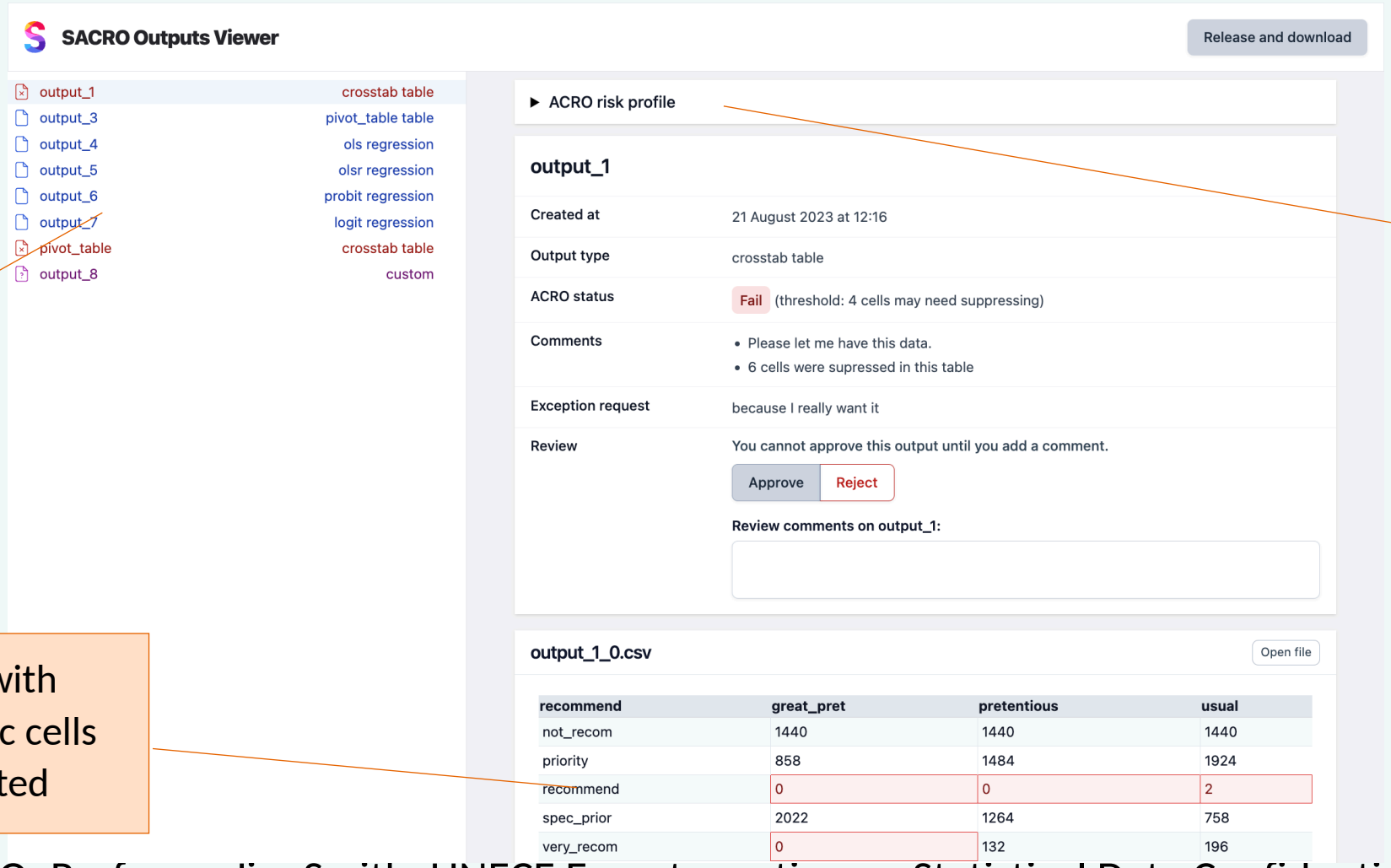
SACRO viewer for TRE output checkers



Outputs to review

Output with problematic cells highlighted

SACRO viewer for TRE output checkers



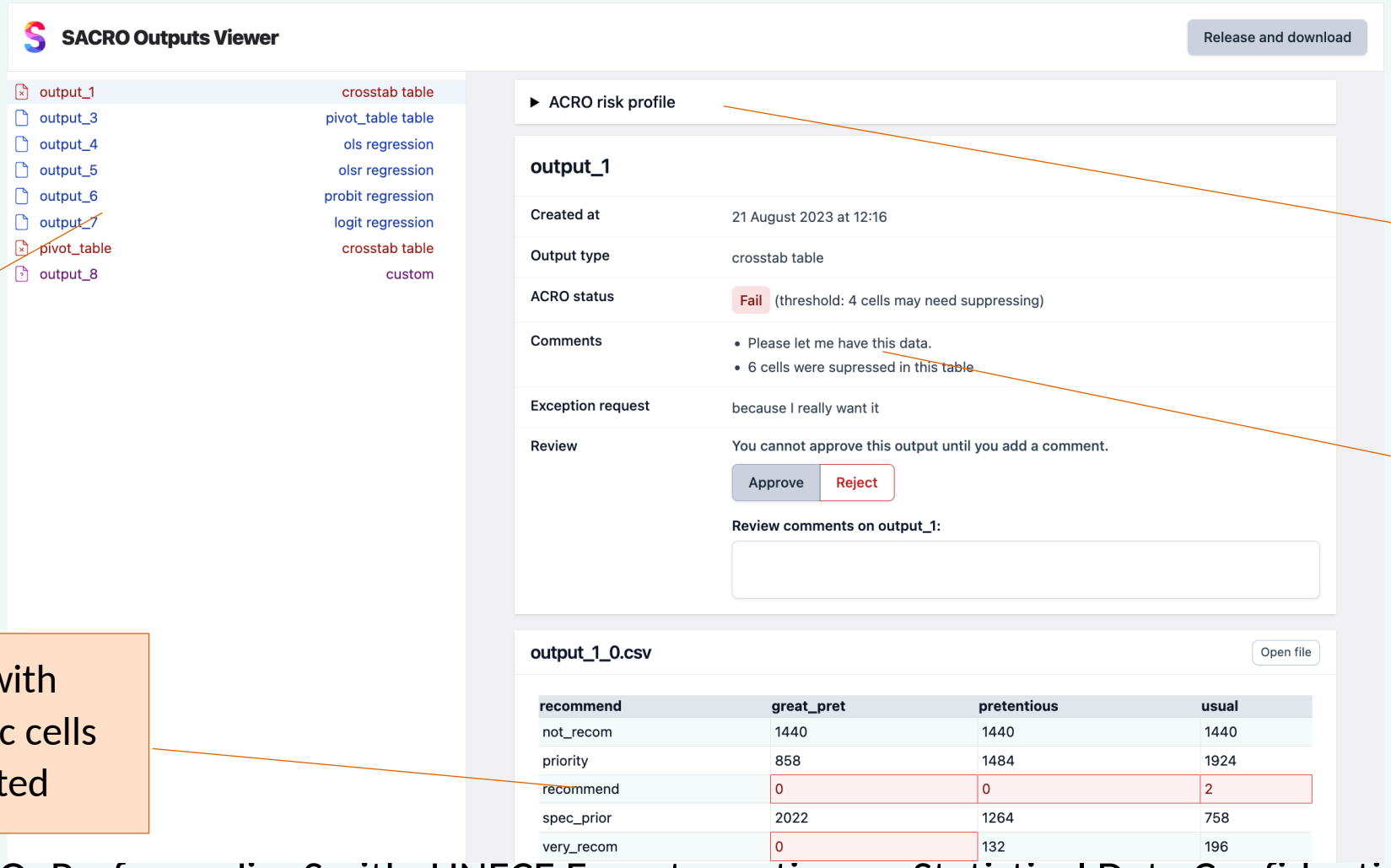
Outputs to review

view TRE risk appetite

Output with problematic cells highlighted

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO viewer for TRE output checkers



Outputs to review

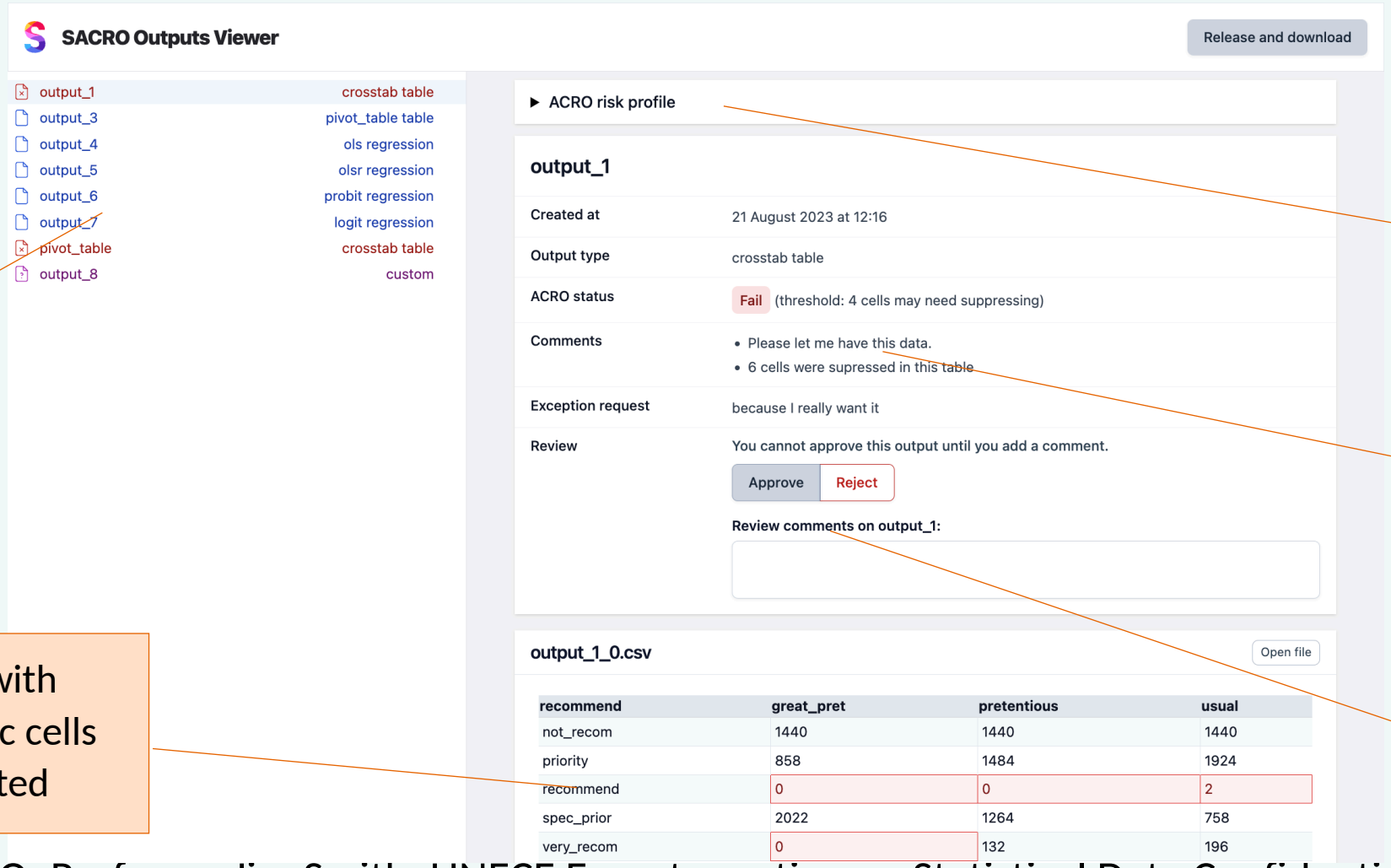
view TRE risk appetite

Type of output
Recommendation
Comments from user
Exception. request

Output with problematic cells highlighted

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

SACRO viewer for TRE output checkers



Outputs to review

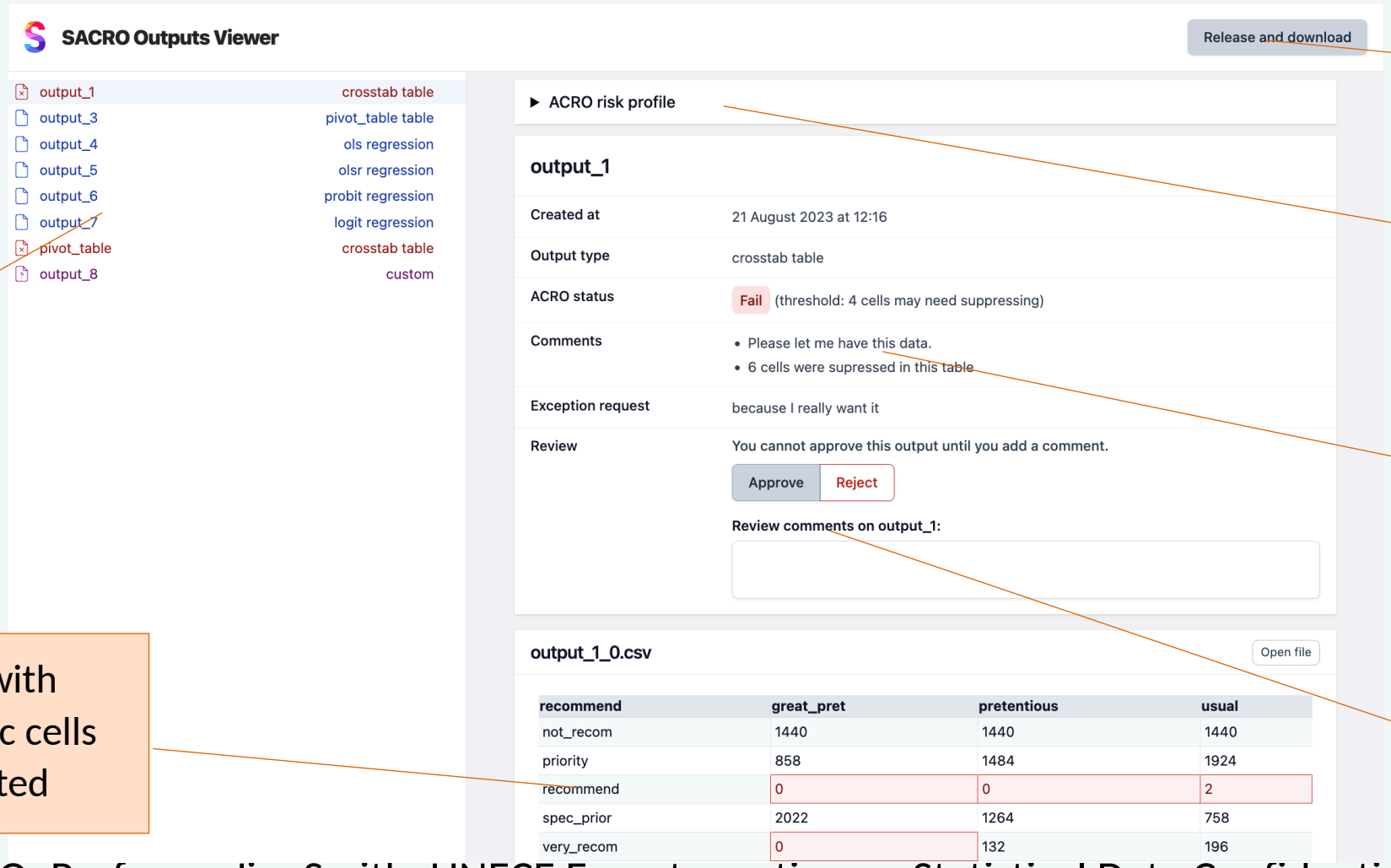
view TRE risk appetite

Type of output
Recommendation
Comments from user
Exception. request

Output with problematic cells highlighted

Comments needed to override recommendation

SACRO viewer for TRE output checkers



Outputs to review

Record overall comments, create release package

view TRE risk appetite

Type of output
Recommendation
Comments from user
Exception. request

Output with problematic cells highlighted

Comments needed to override recommendation

Get Involved

Get involved

Thanks to:

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Get Involved

Get involved

ACRO ‘engine’:

- https://github.com/AI-SDC/ACRO

Thanks to:

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Thanks to:

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening

Get Involved

Get involved

ACRO 'engine':

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening

This work is funded by UK research and Innovation, [Grant Number MC_PC_23006], as part of Phase 1 of the DARE UK (Data and Analytics Research Environments UK) programme, delivered in partnership with Health Data Research UK (HDR UK) and Administrative Data Research UK (ADR UK)